Journal of Behavioral Assessment, Vol. 1, No. 4, 1979

Test-Retest Reliability of Retrospective Self-Reports in Three Populations of Alcohol Abusers

Stephen A. Maisto,[1] Mark B. Sobell,[2] A. Mitch Cooper,[3] and Linda C. Sobell[4]

Accepted November 6, 1979

The reliability of alcohol abusers' retrospective self-reports was evaluated using a time-line follow-back interview technique. Independent groups of male subjects were interviewed in three different treatment settings (outpatient, n = 12; inpatient, n = 12; residential, n = 12) on two separate occasions to assess the test-retest reliability of their self-reports of daily drinking and related events occurring 1 year prior to treatment. Correlational and scatterplot analyses generally showed a high correlation and low discrepancy between the two sets of interview responses for all groups. However, since some population differences were found in the reliability of reports for some variables, further research is needed to delineate the conditions under which reliable and valid retrospective self-reports can be obtained from different populations of alcohol abusers.

KEY WORDS: alcohol abusers; self-reports; reliability; multiple populations.

INTRODUCTION

In contrast to traditional clinical diagnostic and assessment techniques, behavioral assessment methods emphasize objective and verifiable measurement of clients' functioning (Miller, 1973). The alcohol field,

This research was supported, in part, by Grant No. AA07072 from the National Institute on Alcohol Abuse and Alcoholism to Vanderbilt University.

[1]Research Triangle Institute, Research Triangle Park, North Carolina 27709.
[2]Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240.
[3]Community Mental Health Center of Escambia County, Pensacola, Florida 32501.
[4]Dede Wallace Center and Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240.

0164-0305/79/1200-0315$03.00/0

and particularly the area of alcohol treatment outcome evaluation, has been greatly influenced by behavioral assessment procedures. This is exemplified by the recent use of objective measures of drinking behavior, such as in-field breath alcohol assessments (Miller, 1975; Sobell *et al.*, 1979, 1980), the taste test analogue (Marlatt *et al.*, 1973), and tests of acute liver function (Pomerleau and Adkins, 1980; Sobell *et al.*, 1979).

Kazdin and Wilson (1978) recently noted that clinical researchers and practitioners have relied heavily on retrospective self-reports to measure behavior change. Hersen and Bellack (1975) have also identified self-reports as a major component of behavioral assessment procedures. In the alcohol field, the dependence on self-reports is particularly acute, as it is often necessary to obtain retrospective measures of drinking behavior prior to treatment. While baseline measures of drinking behavior are preferable to self-reports, such measurement is usually precluded by ethical and practical considerations. For example, for individuals who risk serious consequences if they continue to drink, it is necessary to use their retrospective reports as the source of baseline data. In these cases, it is essential that the self-reports be reliable and valid.

The present study evaluated the reliability of a recently developed technique designed for use in clinical alcohol treatment research and evaluation. The test-retest reliability of alcohol abusers' self-reports of drinking and related behaviors occurring 1 year pretreatment was assessed. In order to determine the generalizability of the results, three different populations of treated alcohol abusers were studied.

METHOD

Subjects

Three different groups of male alcohol abusers, all of whom met identical screening criteria, participated in the study. Subjects who exhibited any of the following characteristics were excluded from participation: psychiatric problems, mental retardation, organic brain syndromes, or in tratment for longer than 30 days preceding the first scheduled interview date. All subjects volunteered for the study and were assured that their participation would not affect their treatment. Also, at both interviews, all subjects were free of alcohol withdrawal symptoms and were not intoxicated.

Subjects were selected from three alcohol treatment programs located in Nashville, Tennessee. Outpatient subjects (n = 12) were selected from the Alcohol Programs of the Dede Wallace Center, residential treatment

Table I. Descriptive Characteristics of Three Groups of Subjects

Descriptive characteristic		Group		
		Outpatient	Residential	Inpatient
n		12	12	12
Age (yr)	Mean	40.6	39.1	42.0
	Range	30-55	26-61	25-56
Education (yr)	Mean	10.9	9.6	11.3
	Range	4-14	8-13	4-18
Years drinking a problem	Mean	7.9	13.1	13.6
	Range	2-15	1-32	3.5-31
Previous alcohol hospitalizations	Mean	1.25	1.1	4.3
	Median	0.0	0.0	4.5
	Range	0-9	0-4	0-11
Previous public drunkenness arrests	Mean	3.8	36.5	10.1
	Median	0.5	10.5	5.0
	Range	0-33	0-200	0-30
Previous drunk driving arrests	Mean	3.3	1.3	1.7
	Median	2.5	0.5	1.5
	Range	0-8	0-7	0-5
Percent employed		58.3	0.0	0.0
Percent married		75.0	0.0	0.0
Ethnicity: percent white		91.7	100.0	91.7
Previous withdrawal symptoms				
Percent reporting hallucinations		25.0	33.3	41.7
Percent reporting delirium tremens		16.7	16.7	16.7
Percent reporting seizures		8.3	0.0	41.7

subjects ($n = 12$) from the Salvation Army Adult Alcohol Rehabilitation Center, and inpatient subjects ($n = 12$) from the Samaritan Center alcohol treatment program. Descriptive characteristics of the three groups of subjects appear in Table I. These data, gathered at the time of the first interview, clearly demonstrate that the outpatient subjects had greater social resources and less serious alcohol problems than either of the other two groups of subjects.

Procedure

The interview procedures for the three groups were identical, with one exception noted shortly. All subjects were interviewed individually at their respective treatment facility. A standardized questionnaire, read to each subject, included questions about demographic characteristics, drinking history, and daily drinking dispositions for the 360-day period preceding admission to treatment. Subjects were not aware of the categories used to code drinking behavior. Reports of daily drinking were coded into five mutually exclusive categories: days abstinent (no ethanol consumption),

days of limited ethanol intake (any day in which ≤ 3 oz of absolute ethanol was consumed), days of heavy ethanol intake (any day in which > 3 oz of absolute ethanol was consumed), days incarcerated for alcohol-related reasons (jail and hospital incarcerations were combined), and days spent in residential alcohol treatment facilities.

Interviews were conducted by six male and two female interviewers. At the time of the first interview, subjects were not aware that they would be reinterviewed, and in no case did the same interviewer conduct both interviews with the same subject or have knowledge of the subject's previous interview responses. Subjects were informed of the reinterview just prior to its administration. All subjects volunteered to be reinterviewed, gave their informed consent, and were reminded that their participation would not affect their treatment. The same questionnaire was used in both interviews.

The critical feature of the interviews was the gathering of daily drinking behavior data using a time-line follow-back technique (Sobell *et al.*, 1980). This technique involved asking subjects to reconstruct their daily drinking behavior over specified temporal intervals. Subjects were presented with a blank calendar covering the period to be reconstructed and instructed that they were to describe their past drinking as accurately as possible. The drinking information was then gathered using specifically designed interviewing techniques. For example, one interview method involves identifying anchor points, defined as distinct time-bound events (e.g., holidays, weekends, birthdays). In addition to general anchor points, idiosyncratic occurrences were identified, such as days marked by arrests, hospitalizations, illnesses, and entry into treatment. Specifying anchor point events on the calendar facilitates subjects' recall of their drinking on the days of those events, as well as on the days preceding and following those occasions.

A second technique is to identify extended periods of patterned drinking behavior. Subjects are asked to recall the longest series of consecutive days during an interval when absolutely no alcoholic beverages were consumed, as well as the longest number of continuous days of excessive drinking. Similar data are gathered for additional lengthy periods of consecutive abstinent and heavy drinking days, until the subject can report no further distinct episodes, usually periods of less than 7-14 days. Ritualistic drinking episodes, such as weekend or after-work drinking patterns, are recorded in a similar manner.

The one interviewing procedure which differed for the three groups of subjects involved the use of two different test-retest intervals. While a 6-week test-retest interval was used with the outpatient subjects, a 2-week interval was necessary with the other two groups of subjects. Although

equal test-retest intervals would have been preferred, the 6-week interval used for the outpatient subjects was impossible to implement with the other two groups, because a majority of the clients in those programs terminated treatment in less than 6 weeks. Insisting on a 6-week test-retest interval for these two groups would have resulted in high subject attrition as well as unrepresentative samples of the populations, and this situation was viewed as a greater risk to the validity of the study than the use of unequal test-retest intervals. An examination of the two programs' records for the year preceding the study suggested that a representative sample of clients could be obtained by using a 2-week test-retest interval. Finally, because of scheduling constraints, four residential and five inpatient subjects were interviewed using a 3-week interval.

RESULTS

Only two outpatient subjects reported having spent any days in a residential alcohol treatment program. For the year preceding entry into treatment, these subjects reported residential treatment stays of 4 and 3 days and 183 and 184 days, respectively, in their first and second interviews. Since correlational analyses of these data would not have been informative, reliability coefficients for outpatients' days spent in residential treatment facilities were not computed.

Pearson correlation coefficients were computed between the self-reports obtained in the two interviews for all groups of subjects for the daily drinking disposition measures covering the following four cumulative pretreatment time periods: 30 days, 90 days, 180 days, and 360 days. Correlations were also computed for the variables of total years drinking problem, total life hospitalizations, and total life arrests. These test-retest reliability coefficients are presented in Table II. Except for a few of the drinking behavior variables for the residential and inpatient subjects, the correlation coefficients indicate a high degree of reliability for the self-reports of all three groups of subjects. The lowest reliability coefficients for residential subjects were for numbers of days abstinent and number of days of heavy drinking, whereas inpatient subjects showed the least consistency for self-reports of number of days of limited drinking and number of days of heavy drinking.

Scatterplots comparing subjects' first and second interview responses for drinking dispositions during the 360-day pretreatment interval and drinking problem history questions are shown in Fig. 1, 2, and 3. Only data for the 360-day interval are presented for purposes of brevity, and because in most cases those data reflect the general pattern of results at shorter inter-

Table II. Test-Retest Reliability Coefficients for Self-Reports of Drinking and Related Behaviors Occurring 30, 90, 180, and 360 Days Before Treatment in Three Different Populations of Alcoholics[a,b]

	Time interval											
	30 days			90 days			180 days			360 days		
Variable	Opt.	Res.	Inpt.	Opt.	Res.	Inpt.	Opt.	Res.	Inpt.	Opt.	Res.	Inpt.
Days abstinent	0.79[c]	0.69[c]	0.88[c]	0.85[c]	0.62[d]	0.94[c]	0.92[c]	0.61[d]	0.98[c]	0.93[c]	0.33	0.98[c]
Days ≤ 3 oz ethyl alcohol	0.92[c]	—	−0.13	0.90[c]	0.67[c]	−0.13	0.92[c]	0.23	−0.15	0.94[c]	0.88[c]	−0.12
Days > 3 oz ethyl alcohol	0.88[c]	0.62[d]	0.50[d]	0.91[c]	0.43	0.60[d]	0.89[c]	0.54[d]	0.52[d]	0.94[c]	0.40	0.73[c]
Days incarcerated	0.98[c]	0.92[c]	0.77[c]	0.89[c]	0.996[c]	0.93	0.98[c]	0.95[c]	0.84[c]	0.98[c]	0.97[c]	0.96[c]
Days residential	—	0.91[c]	0.96[c]	—	0.88[c]	0.90[c]	—	0.87[c]	0.95[c]	—	0.86[c]	0.91[c]

	Outpatient	Residential	Inpatient
Years drinking a problem	0.82[c]	0.74[c]	0.87[c]
Total life hospitalizations, alcohol related	0.98[c]	0.81[c]	0.88[c]
Total life arrests, alcohol related	0.97[c]	0.98[c]	0.97[c]

[a]Opt., Outpatient (n = 12); Res., residential (n = 12); Inpt., inpatient (n = 12).

[b]Correlations are not reported for (1) the outpatients because they reported too few days of residential stays to compute meaningful correlations and (2) the residential clients as no subject reported days of ≤ 3 oz of alcohol consumption during 30 days pretreatment in either interval.

[c]$p < 0.01$ for one-tailed tests.

[d]$p < 0.05$, for one-tailed tests.

vals.[5] The scatterplots reveal several interesting differences among the three groups in reports of both drinking history and drinking dispositions. In this regard, the outpatients tended to report a longer history of drinking problems during the second interview compared to the first interview. However, both the residential and inpatient groups reported a shorter history of drinking problems in the second interview compared to the first interview. The residential group also reported more alcohol-related arrests in the second interview.

The scatterplots of the drinking disposition data (Figs. 1 and 2) indicate that the absolute magnitudes of discrepancies were relatively small. While the outpatients generally reported fewer abstinent days, fewer heavy drinking days, fewer incarcerated days, and more limited drinking days during the second interview as compared to the first interview, the discrepancy patterns for the other two groups were somewhat different.

The residential subjects generally reported fewer abstinent days in the second interview; however, those residential subjects who reported limited drinking in the first interview tended to report more limited drinking days in the second interview. Residential subjects also tended to report more heavy drinking days, fewer days incarcerated, and more residential treatment days during the second interview.

Figures 1 and 2 also suggest that the inpatient subjects' self-reports of abstinent days showed no systematic trends across interviews. Furthermore, like the residential subjects, the inpatients generally reported a low frequency of limited drinking days. Once again, these reports were more frequent in the second interview (one subject did, however, show a marked shift). Inpatient subjects tended to report fewer heavy drinking days in the second interview. They also showed no systematic trends in discrepancies over interviews for days incarcerated; however, they did report that they spent fewer total days in residential treatment facilities as compared to subjects in the residential group.

The discrepancy data were further analyzed by examining the relationship between the frequency (number of days) with which a behavior was reported in the first interview and the number of those days that were reported differently in the second interview—"behavior change days." Since it could be argued that part of the high degree of consistency seen in Figs. 1, 2, and 3 may be attributable to the low frequency of some of the

[5]The exception to this was in the residential subjects' data. Several subjects in this group showed major shifts in their reports of abstinent, limited drinking, and heavy drinking days. Most of these shifts pertained to the first 6 months of the year preceding treatment, resulting in variability in the magnitude of the reliability coefficients for reports of the drinking behaviors in the relevant two intervals.

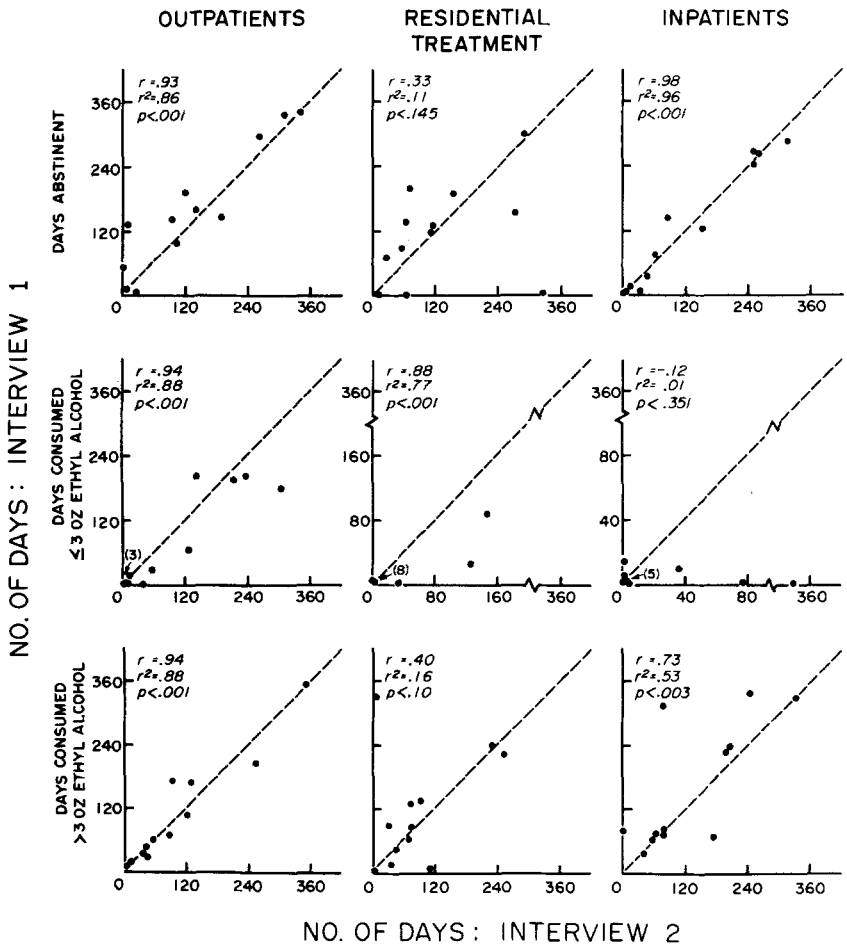

Fig. 1. Scatterplots of the first and second interview responses for number of days abstinent, number of days of limited (≤ 3 oz ethanol) drinking, and number of heavy (> 3 oz ethanol) drinking days occurring 360 days immediately preceding treatment admission. Outpatient, residential, and inpatient groups each contained 12 subjects.

behaviors, Pearson correlations were computed for all 36 subjects, and for each of the three groups. For the total sample, two significant ($\alpha = 0.05$, two-tailed tests) correlations were obtained: number of heavy drinking days ($r = 0.46$) and number of days spent in a residential facility ($r = 0.47$). Both the outpatient and inpatient groups revealed only one significant correlation—residential days, $r = 0.69$ and $r = 0.88$, respectively; the residential group, however, yielded significant correlations for number of limited drinking days ($r = 0.64$), number of days of heavy ethanol consumption ($r = 0.61$), a number of days incarcerated ($r = 0.93$).

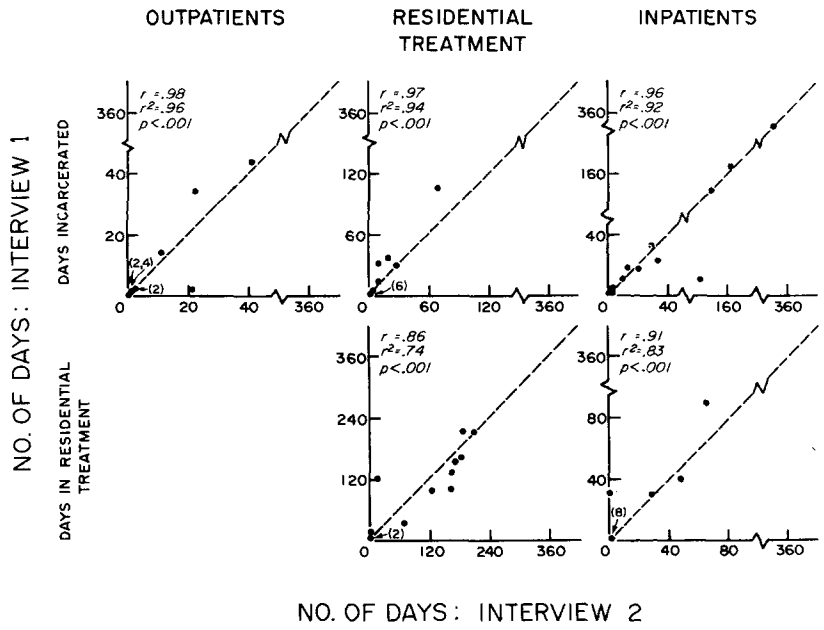

Fig. 2. Scatterplots of the first and second interview responses for number of days spent incarcerated in a jail or hospital for alcohol-related reasons and number of days spent in residential treatment during the 360 days immediately preceding treatment admission. A plot was not drawn for the outpatient sample's residential treatment reports because this event occurred for only two subjects during the year preceding their admission to treatment (see text). Outpatient, residential, and inpatient groups each contained 12 subjects.

To summarize, with the few exceptions noted earlier, the correlational analyses suggest uniformly high reliability for the outpatient subjects and generally high reliability for the residential and inpatient subjects. Furthermore, the discrepancies in self-reports in most cases were not strongly related to the frequency with which those behaviors were reported to have occurred in the first interview.

A final set of analyses was performed to identify factors that predicted discrepancies in subjects' self-reports. To this end, the number of behavior change days was collapsed across the five drinking dispositions for the 360-day interval and divided in half (because of the dependencies inherent in the measures). This variable was regressed on four subject characteristics reported in the first interview: (1) highest occupational level achieved (Hollingshead Scale), (2) reported history of alcohol withdrawal symptoms (hallucinations, delirium tremens, or seizures), (3) years of education, and (4) years drinking was reported as a problem. The subject's group (treatment program) was also entered into the equation. All subjects were included in this analysis; a forward stepwise regression procedure was used,

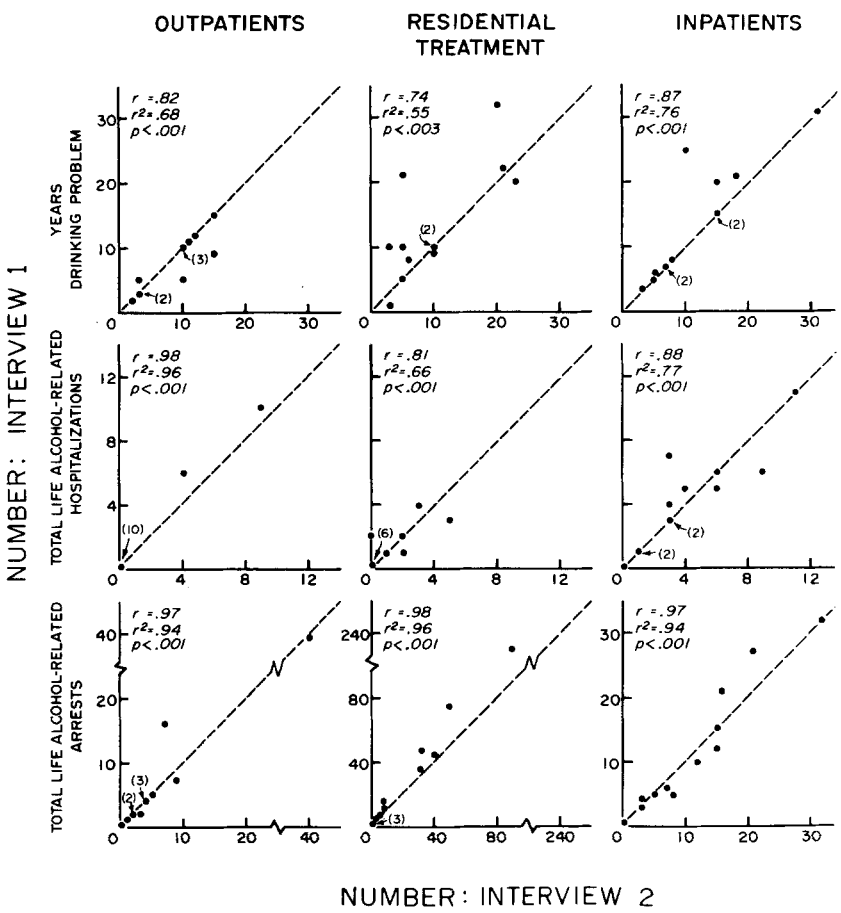

Fig. 3. Scatterplots of the first and second interview responses for years drinking a problem, total life alcohol-related hospitalizations, and total life alcohol-related arrests (sum of reports of the number of arrests for driving under the influence of alcohol and drunk in public). Outpatient, residential, and inpatient groups each contained 12 subjects.

with liberal criteria for entering a variable into the equation (all of the variables met the criteria). The analysis revealed that these variables did not significantly predict the total discrepancy score, $R = 0.51$, $F(6, 29) = 1.70$, $p > 0.05$. The two factors that accounted for the most variance were occupational level (7.8% of 26%) and treatment group designation (6.1%). The variable that contributed least to the variance was number of years of education (2.4%).

DISCUSSION

Since self-reports of drinking behavior are central to assessment and treatment evaluation procedures, it is imperative to obtain highly reliable and valid self-reports of such behaviors. Using a time-line follow-back interview technique, this study found that three different groups of alcohol abusers with differing drinking histories gave relatively reliable retrospective self-reports of their daily drinking and related behaviors for the year preceding treatment. While most correlations were fairly high, low correlations occurred for the residential and inpatient groups on several of the drinking behavior variables. These findings are particularly important, because the baseline rate of a behavior frequently determines how much significance is accorded to changes in the rate of that behavior during and following treatment.

The analyses of the discrepancies between the first and second interview reports revealed a few systematic shifts in each of the three groups for different drinking behavior variables. Although the patterns in the discrepancy data are interesting, further investigations are needed to interpret their meaning. The supplementary analyses of the discrepancy data for the entire pretreatment year generally showed that the number of behavior change days for a drinking behavior variable was not strongly correlated with the frequency with which that behavior was reported to have occurred in the first interview. Furthermore, regression analyses revealed that neither subject characteristics nor treatment program designation significantly predicted the total number of behavior change days for the pretreatment period. These findings suggest that future research should investigate other variables (e.g., interviewer and situational characteristics) which could produce discrepancies in self-reports.

In considering this study, it must be remembered that the results were obtained using highly restricted interviewing conditions: (1) subjects were not intoxicated when interviewed, (2) interviews took place at a treatment facility, and (3) interviews were conducted by trained interviewers using a time-line follow-back interview procedure. Conducting interviews in different settings (i.e., in field follow-up), using different methods or interviewing for different purposes (e.g., job interviews) could produce less reliable self-reports. This study also demonstrated that the reliability of alcohol abusers' reports of certain alcohol-related behaviors varies with treatment population. On the basis of these data, it is imperative that further research be conducted to develop procedures and delineate conditions under which reliable and valid self-reports can be obtained from diverse populations of alcohol abusers.

REFERENCES

Hersen, M., and Bellack, A. S. *Behavioral Assessment: A Practical Handbook.* New York: Pergamon Press, 1976.

Kazdin, A. E., and Wilson, G. T. Criteria for evaluating psychotherapy. *Archives of General Psychiatry,* 1978, *35,* 407-416.

Marlatt, G. A., Demming, B., and Reid, J. B. Loss of control drinking in alcoholics: An experimental analogue. *Journal of Abnormal Psychology,* 1973, *81,* 233-241.

Miller, P. M. Behavioral assessment in alcoholism research and treatment: Current techniques. *The International Journal of the Addictions,* 1973, *8,* 831-837.

Miller, P. M. A behavioral intervention program for chronic public drunkenness offenders. *Archives of General Psychiatry,* 1975, *32,* 915-918.

Pomerleau, O., and Adkins, D. Evaluating behavioral and traditional treatment for problem drinkers. In L. C. Sobell, M. B. Sobell, and E. Ward, (Eds.), *Evaluating Alcohol and Drug Abuse Treatment Effectiveness: Recent Advances.* New York: Pergamon Press, 1980.

Sobell, M. B., Maisto, S. A., Sobell, L. C., Cooper, A. M., Cooper, T., and Sanders, B. Developing a prototype for evaluating alcohol treatment effectiveness. In L. C. Sobell, M. B. Sobell, and E. Ward (Eds.), *Evaluating Alcohol and Drug Abuse Treatment Effectiveness: Recent Advances.* New York: Pergamon Press, 1980.

Sobell, M. B., Sobell, L. C., and VanderSpek, R. Relationship between clinical judgment, self-report and breath analysis measures of intoxication in alcoholics. *Journal of Consulting and Clinical Psychology,* 1979, *47,* 204-206.